# STANDARD OPERATING PROCEDURE FOR DATABASE OPERATIONS

Environmental Health and Safety Division
Research Triangle Institute, Research Triangle Park, North Carolina

Prepared by: ____________ Date: Aug 12 2003

Reviewed by: ____________ Date: 8-15-03

Approved by: ____________ Date: 8-15-03

This document contains confidential information and should not, in whole or part, be further released, distributed, or disclosed beyond the recipient of the document.

**CONTENTS**

## CONTENTS (continued)

# STANDARD OPERATING PROCEDURE FOR DATABASE OPERATIONS

## 1.0 Introduction

### 1.1 Scope

These operating procedures cover database operation activities performed by program data processing staff. Data entry activities, such as sample handling analysis laboratory (SHAL) sample processing, are included in the relevant standard operating procedures (SOPs) for the group or activity that performs the data entry.

### 1.2 Requirements

Most of these procedures assume a familiarity with general database concepts and use of MS Access and SQL Server programming tools (such as the Query Analyzer and Enterprise Manager). General Windows NT Server management skills are also assumed for supervisory personnel.

### 1.3 Hardware/Software Environment

Internal Server - RTI maintains an internal database server for use with the speciation project. This server runs Microsoft SQL Server version 7 on the Microsoft Windows NT Server operating system. Only internal RTI personnel are allowed access to this internal server (individual accounts are set up as described below).

External Server - An external server (outside RTI's Internet firewall) is used to post data for EPA and state review. This machine holds the final reports, only.

NOTE: The names of specific forms, queries, reports and programs to be run are italicized throughout this SOP.

## 2.0 Create Data Contact Account on External Server

2.1 Summary of Task

This procedure describes the steps necessary to create a data contact account on the external server (GEOS1.rti.org). This procedure requires that the user have administrative rights on the external server.

2.2 Procedure

2.2.1 Use User Manager and select domain of geos1.rti.org.

2.2.2 Add user account and set password. Set password to not expire, not change on initial login, and not be changed by user.

2.2.3 Add user to PM25 group.

2.2.4 Create directory for user's data under correct EPA directory.

2.2.5 Set security for directory. Grant read/list access to user, Delivery Order Project Officer (DOPO) and overall project officer (EPA01).

2.2.6 Send account, directory, and password to appropriate DOPO for transfer to data contact.

## 3.0 Create NT Domain Account for New Temporary Employee

3.1 Summary of Task

This is done for each new temporary service employee working in SHAL. RTI employee domain accounts are created by their center office as part of the hiring procedure. This procedure requires administrative rights on the internal server.

3.2 Procedure

3.2.1 Laboratory Supervisor goes to ITS web site and completes an *account request form*. Request ONLY NT Domain Account. Be certain to mark employee as temporary on the form.

3.2.2 ITS returns (rejects) account form to laboratory supervisor, requiring center director approval.

3.2.3 Laboratory Supervisor forwards returned e-mail to center director for approval

3.2.4 Center Director approves new account and forwards approval to ITS.

3.2.5 ITS creates new account. Notifies Laboratory Supervisor by e-mail.

3.2.6 Laboratory Supervisor forwards account information to the Center Information Management Systems (IMS) Supervisor and Database Supervisor.

3.2.7 Database Supervisor creates SQL server account and grants appropriate disk access on server. Notifies Laboratory Supervisor that access has been granted.

## 4.0 Delete NT Domain Account for Terminated Temporary Employee

4.1 Summary of Task

The laboratory supervisor must notify ITS whenever a temporary employee is no longer needs access so that their NT Domain Account will be closed.

4.2 Procedure

4.2.1 Notify ITS, CEM IMS Supervisor, and Database Supervisor of the appropriate account to be deleted.

4.2.2 Remove account from file access and deletes SQL Server account.

## 5.0 Add RTI employee to SHAL database users group

5.1 Summary of Task

This is done for RTI employees who need database access. This procedure requires administrative and SQL administrative rights on the internal server.

5.2 Procedure

5.2.1 Notify Database Supervisor of person to be granted SHAL database access.

5.2.2 Laboratory Supervisor adds account to File Access Group and creates SQL Server Account. Assigns database access and appropriate database roles to SQL server account.

## 6.0 Remove RTI Employee from SHAL Database Users Group

6.1 Summary of Task

This is done whenever database access is no longer needed for an RTI employee. This procedure requires administrative and SQL administrative rights on the internal server.

6.2 Procedure

6.2.1 Laboratory Supervisor notifies Database Supervisor that database access is no longer needed.

6.2.2 Database Supervisor removes account from access group and deletes SQL Server login account.

## 7.0 Process Delivery Order and Schedule Associated Sampling and Analysis Events

7.1 Summary of Task

This procedure describes the operations necessary to process an incoming delivery order and to schedule the associated analytical and sampling requests.

7.2 Procedure

7.2.1 Get delivery order information from EPA DOPO.

7.2.2 Determine information needed for delivery order processing from information provided by EPA DOPO.

7.2.3 Enter information for delivery order into database.

7.2.4 Run *delivery order form* (as report) from database.

7.2.5 Prepare folder for delivery order paperwork. Place delivery order form into folder.

7.2.6 Identify each site on delivery order and determine if it is on RTI’s list of sites with sampler and AIRS information.

7.2.7 Determine if the site is listed on RTI's list of sites with sampler and AIRS information.

7.2.8 Enter sampler and AIRS information for new site into database, if necessary.

7.2.9 Determine sampler type and analysis list needed for each site listed on the delivery order. (This assumes that the same type is required throughout the delivery order.)

7.2.10 Use lookup list to determine the sampling configuration needed for the selected sampling type and analysis list.

7.2.11 Determine beginning and end dates for each site. Use measurement request generation program to create measurement requests for each site and date combination.

7.2.12 Print sampling request forms for location and file in processing folder(s).

7.2.13 Review sampling forms to verify that scheduling is correct.

7.2.14 After all samples have been scheduled, set delivery order status to requests scheduled.

7.2.15 Select delivery order for proofing.

7.2.16 Compare summary report to delivery order - make changes, if necessary.

7.2.17 Update SHAL schedule calendar to reflect additional work load.

## 8.0 Receive Data from Laboratory

8.1 Summary of Task

This step describes receipt of data (in spreadsheets) from the analysis laboratories for direct import into the database.

8.2 Procedure

8.2.1 Receive spreadsheet containing analytical results from laboratory.

8.2.2 Move spreadsheet data onto server into the appropriate laboratory file.

8.2.3 Review laboratory files to make sure that they are in the correct format for import. Make corrections to format as necessary for automatic import into database.

8.2.4 From the database use the *import analytical data* form to automatically import analytical results into the database.

8.2.5 If errors occur during import, do not commit the transaction, identify and correct any problems with analytical data before importing the data.

8.2.6 Move imported laboratory results files into the added to database folder within each laboratory folder on the server.

## 9.0 Prepare Monthly Analytical Data Report

9.1 Summary of Task

This procedure describes the preparation of the analytical data report, which is sent to the EPA DOPO each month.

9.2 Procedure

9.2.1 Perform preliminary duplicates check - by running the *DignoseDuplicateRows* program.

9.2.2 Correct any duplicates, as necessary.

9.2.3 Make a copy of the main database for use in report checking by running the *Transfer to QC Draft* program using the SQL Server Data Transformation Services.

9.2.4 Select last sample for delivery date by setting the correct value for the last sampling date by editing the date in the *ForceApproveTestBatch.sql* program.

9.2.5 Batch approve samples by running the SQL script by running the *ForceApproveTestBatch.sql* program.

9.2.6 Fix problems with reported uncompleted samples (in main database). Rerun everything to this point if any unaccounted sampling events remained.

9.2.7 Generate Analytical Report Data in draft database by running the program *Export_Report_Data.*

The Export_Report_Data procedure performs the following calculations:

1) Ambient concentration = analyte mass /sampler volume (for appropriate sampler channel), where sampler volume is from Field Custody Chain of Sampling Form. If no sampler volume was supplied but an average flow and elapsed time were supplied, then sampler volume = average flow * elapsed time.

2) Sample concentration uncertainty = sample mass uncertainty / sampler volume (for appropriate sampler channel)

3) Sample mass uncertainty = greater of

a) lab blank uncertainty and

b) square root of (s_analytical$^2$ + (M * sr_volume)$^2$)

where:

s_analytical = laborarory analysis uncertainty,

M = sample mass

sr_volume = volume uncertainty (relative) for appropriate sampler channel

Notes:

1) Concentrations and concentration uncertainties are not defined for field and trip blanks as there is no sample volume.

2) The sampler volume uncertainty (sr_volume) is currently assummed to be ± 5% for all sampler

9.2.8 Check report data for duplicate rows by running the program *Check_for_Dups* with the argument of the latest batch number. Correct any problems and regenerate all steps to this part, if necessary.

9.2.9 Open the report generation program (Report.mdb) and verify that its tables are linked to the database copy used for QA review. Relink to the correct database with linked table manager, if necessary.

9.2.10 Run the report program to generate draft copies of the output reports.

9.2.11 Notify QA Officer that draft reports are ready for review.

9.2.12 After QA approval, repeat procedures previously completed on copy of database on the main database.. Copy final copy of reports to external web site.

9.2.13 Remove any old reports from the web site.

9.2.14 Notify project officer that reports are ready for review.

9.2.15 Notify DOPOs that reports are ready.

9.2.16 Make CD copies of web site for EPA project officer and the master project file.

9.2.17 Deliver CD to EPA project officer.

## 10.0 Prepare results for AIRS

10.1 Summary of Task

This procedure describes the preparation of the monthly AIRS data report.

10.2 Procedure

10.2.1 Copy current database over training db in preparation for trial generation and review.

10.2.2 Run the Program *Fill_AIRS_table* with argument of the batch to be delivered - this generates all data in the batch.

10.2.3 Using the main data entry application, set the event's delivery status to 5 - 'AIRS reprocessing needed' for all events that need to be regenerated due to state reviews, etc.

10.2.4 Remove any data that needs reprocessing from the internal AIRS processing table by running the program *Delete_challenged_data* with argument of the AIRS batch to be delivered,.

10.2.5 Perform any needed recalculations by running the program *Update_challenged_data* with the argument of the AIRS Batch to be delivered.

10.2.6 Check for any duplicate records by running the program *Check_for_dups* with argument of the batch to be delivered--fix any duplicates found.

10.2.7 Now that all records have been reprocessed, the AIRS output file is generated by using the menu in the *AIRSApp.adp* program

10.2.8 Import AIRS file into the *AirsFIle* table in *AIRS_File_Checks.mdb* using the custom import/export specifications that are saved in the database.

10.2.9 Run QA procedures, as described in AIRS QA Procedures, below.

10.2.10 correct any errors and restart process up to this step - make note of any changes on QA review form.

10.2.11 Send file to QA officer for review of AIRS data. Make any necessary corrections.

10.2.12 QA officer reviews and approves file.

10.2.13 Rerun procedures on main database.

10.2.14 Update the delivery status by running the *SetAIRSDeliveryStatus* program with argument of the AIRS Batch to be delivered.

10.2.15 Submit each sub file to AIRS AQS system - see EPA's AIRS documentation for procedures.

10.2.16 Subdivide approved AIRS file into subfiles (typically 6) to get files small enough for posting in AIRS.

10.2.17 Notify EPA that results have been posted.

## 11.0 AIRS QA Procedures

11.1 Summary of Task

This procedure describes the steps needed to review and approve AIRS data before posting to AIRS/AQS.

11.2 Procedure

11.2.1 Run the AIRS file - Make a copy of the current database for use in QA review.

11.2.2 Generate a draft AIRS file using the *AIRSApp.adp* database application. Make sure the application is connected to the database copy before file generation.

11.2.3 Once the AIRS file has been generated, import the file into the *AIRSFile* table in the *AIRS_file_checks.mdb* QA software using the custom import/export specifications contained in this software program.

11.2.4 For the first run of the AIRS file to check for errors inherent in the data. See Appendix A, Quality Control Summary Sheet. Query names for each check are in italics.

Perform the first 4 four steps in the AIRS Batch File Quality Control Summary Sheet (Appendix A) to check for duplicates, modify records, new sites, and null start hours. If there are any new sites, check for complete monitor information. If the monitor information is incomplete generate pdf files showing the monitor information for transmittal through DOPOs to request information from the appropriate agencies. Once all monitor information is received, post monitor records in AIRS. Also add the AIRS code, POC and expected record count to the *Expected_AIRS_Counts* table in the *AIRS_file_checks.mdb* database. Once the new sites are added to the expected counts table, the *Check_Record_Count2* query can be run to check the record counts. Also update the *Check_StartHour1* query batch number to the current batch number, then run the *Check_StartHour3* query to check for odd start times.

Any errors found at this point must be corrected in the main database and marked for reprocessing. If events with duplicate dates cannot be rectified by reviewing the data in the database, send off a request to appropriate SHAL personnel so that they can review the hard copy paperwork and send inquiries to the site operator if necessary.

11.2.5 Subsequent Runs of AIRS File - After all of the corrections have been made from the first run of the AIRS file, the AIRS file is run again to verify that all of the changes have been made correctly as well as checking that all of the state requested changes have been made correctly.

To check the state requested changes, first run the *Temp_Trog13* query on the main database to get a list of all of the events in the current batch which are marked for reprocessing. Check in the appropriate batch folder for copied request change e-mails. Also check for any other e-mails with changes in personal inbox related to data changes. Make sure that all events marked for reprocessing have a requested change associated with them as well as verify that all e-mails with requested changes have the events marked for reprocessing in the database. If anything is missing, contact person responsible for making the changes. If there is any question on the interpretation of changes, contact the Quality Assurance (QA) officer.

Once all of events marked for reprocessing are verified, copy over the database and follow the procedures to generate an updated AIRS file and import it into the *AIRS_file_checks.mdb* QA database. Run all of the queries previously run to verify that all errors have been corrected. Because of the large datasets, the remaining queries are run on smaller static batch-specific tables, but these must be refreshed for each report batch. To do this, run the *Delete_New_AIRS_Data* and *Delete_New_Report_Data* queries, to delete the data from the previous batch. Now refill the tables with the current batch's data by updating the batch number to the current batch number in the *Convert_Report_Data1* query, then running the *Append_New_AIRS_Data* and *Append_New_Report_Data* queries, to append the new data. Now all of the queries specified on the Quality Control Summary Sheet (Appendix A) should be run to check the AIRS file and verify all of the changes. Note that *Check_Delivered_Records1* query, requires that the data from the previous AIRS batch be appended to the *AIRS_Posted* table using the *Append_AIRS_Posted* query. This will verify that no events are in the AIRS file that were in a previous AIRS file.

If errors still remain in the AIRS file, corrections should be made in the main database, and the AIRS file reprocessed. Any manual changes to the AIRS file as well as any data withheld from the AIRS file should be noted in the Quality Control Summary Sheet.

Once the final AIRS file is generated, a completed Quality Control Summary Sheet, the printout results of all of the queries, and printouts of all state requested changes are submitted to the project QA officer for review. Once the QA officer has approved the QC packet, the file can be loaded into the EPA's AQS. Typically the AIRS file is split into 6 separate files of approximately 15,000 records in order to minimize loads on the AQS.

11.2.6 After all of the AIRS data has been posted to AIRS, run the AIRS processing on the main database and mark events as delivered to AIRS.

## 12.0 Database Backup

12.1 Summary of Task

Backup and restore procedures are covered in the Speciation Program Disaster Recovery Plan. Two types of backups are performed:

- backup of SQL Server database to backup and transaction files,
- automated nightly tape backup of all disk space (including SQL Server backup files).

## 12.2 SQL Server Backup

A full SQL Server backup file is generated weekly and written to tape as part of the automated nightly tape backup job. SQL Server transaction logs are generated each hour during the business week (Monday to Friday from 7 am to 5 pm). The transaction logs, in combination with the full backup, permit restoration of the entire database with all data that was present at the time of the transaction log creation. Backup files and transaction logs are copied each hour (on the half hour) to disk space on a second server in a different building for safekeeping.

## 12.3 Nightly Tape Backup

Tape backups are run each night, with a full backup done weekly and incremental backups done each night (except for the night when the full backup is run). Tapes are changed weekly and the most recent week's backup file is stored in different building from the server.

## APPENDIX A
## PM 2.5 Speciation AIRS Batch File Processing Quality Control Summary Sheet for Report Batch ___

**Basic AIRS file checks (required for AIRS update)**

1. Check that no duplicate rows are in AIRS file by running query *Check_Duplicates1*.

2. Check that all modified records are appropriate by running query *Check_Modifies1*. Explain any re-deliveries of AIRS events or manual modification to correct modifies

3. Check that Null start hours only occur for null value codes/decimal position is null by running query *Check_Start_Hour1*.

4. Check for addition of new locations or modifications of current locations by running query *Check_New_Sites3*. Explain any changes:

5. Check that there are no records delivered in previous batches for the same date that are not modified records by running the query *Check_Delivered_Records1*.

6. Explain any problems found during AIRS edits which require manual edits in final AIRS file.

**AIRS file Checks against Original Report Batch Data**

1. Check for odd start times, by enter current batch number in the query *Check_StartHour1*, then running the *Check_StartHour3* query. Explain any odd times found.

2. Check for records in AIRS file not in original report, by running the if any, *Check_Not_in_Report1* query. Explain any such records.

3. Check for any records in original report not in the AIRS file by running the *Check_Not_in_Airs1* query. Explain any such records.

4. Check for duplicates in original report by running the query *Check_Report_Duplicates* (used only to rectify difference in counts,

5. Rectify the number of records in the AIRS file vs. the original report to verify that all appropriate records in the report made it into the AIRS file and no spurious records were created in the AIRS file that were not in the original report.

6. Compare record counts in the AIRS file vs. the expected record counts for each AIRS code/POC by running the *Check_Record_Count2* query. Explain any counts that do not match.

7. Compare values and AIRS null values codes in the AIRS file to those in the original report. Check 100% of those with an absolute difference of >=1. Spot check those that are different by 10% or more, but with an absolute difference <1. Spot check any differences between 1% and 10% to make sure that differences are only caused by rounding. Explain any differences not caused by rounding. This may be checked by using the *Check_Reprocessing* query (change criteria as necessary to review different ranges of differences)

8. Compare AIRS validity codes in the AIRS file to those in the original report by running the *Check_Validity query*. Explain any differences found.

**Verification of AIRS reprocessing changes**

1. Summarize changes made at the request of the states (check with person responsible for making the changes) and verification that these changes were updated with reprocessing.

    Check in appropriate batch folder in for copied e-mails. Also check for any other e-mails with changes in personal inbox.

2. Summarize changes made in response to problems found during the AIRS processing and verification that these changes were updated with reprocessing.

**Events in this report batch that were withheld from update in AIRS:**

Explain reasons for any data withheld.